Tópico 06 - Memória externa

Hugo Vinícius Leão e Silva

hugovlsilva@gmail.com, hugo.vinicius.16@gmail.com, hugovinicius@ifg.edu.br

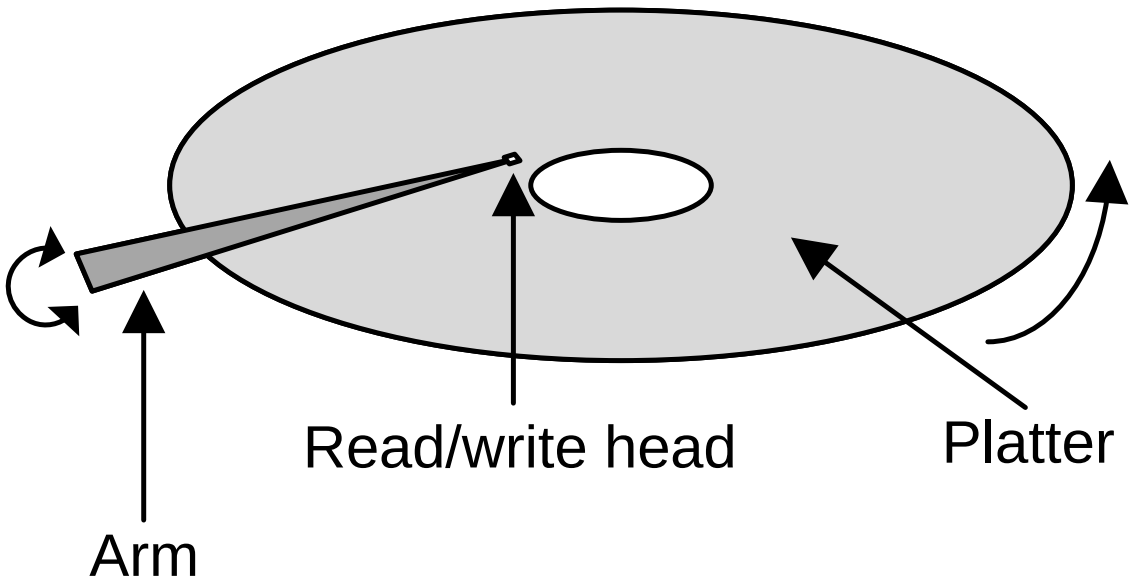
Instituto Federal de Educação, Ciência e Tecnologia de Goiás
Campus Anápolis
Curso de Bacharelado em Ciência da Computação

22 de julho de 2021

Visão geral

Disco magnético

Figura: Disco magnético

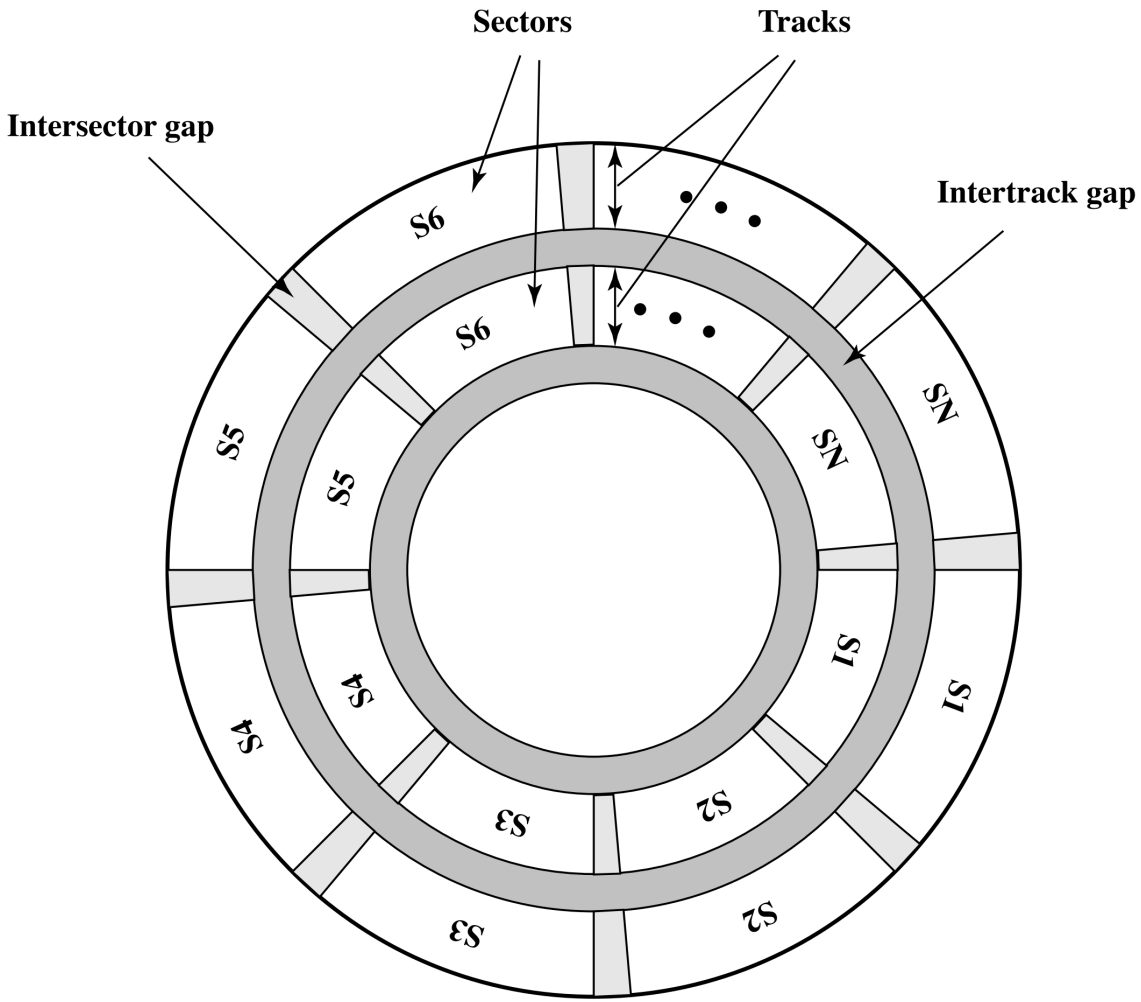


Disco magnético

- O disco magnético é construído de material não-magnético – substrato – revestido com material magnetizável;
- Substrato pode ser alumínio, liga de alumínio ou vidro;
- É formado por círculos concêntricos – **trilhas** – da mesma largura que a cabeça de leitura/gravação;
- Cada trilha é formada por **setores** onde são gravados os dados – normalmente 512 ou 4096 bytes (*Advanced Format*);
- Há espaços entre trilhas e setores para aumentar as tolerâncias, diminuir erros por *misalignments* e por interferência de campos magnéticos.

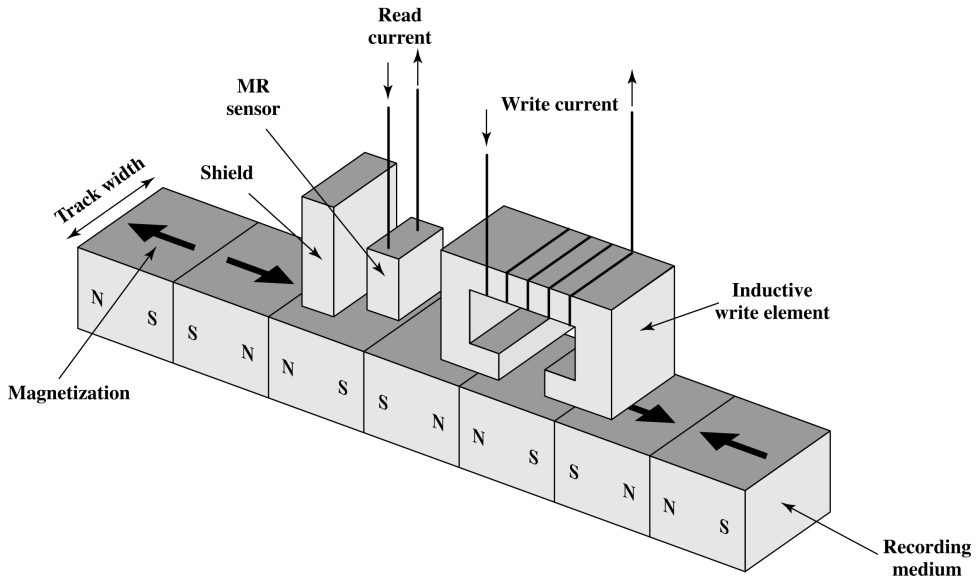
Disco magnético

Figura: *Layout* de um disco magnético



Disco magnético

- A cabeça de leitura/gravação é feita de um núcleo ferromagnético e seu funcionamento baseia-se:
  - Lei de Faraday: uma variação no campo magnético produz uma tensão elétrica em uma bobina próxima. A leitura é feita por um sensor magneto-resistivo, onde é induzida uma corrente elétrica;
  - Lei de Ampère: a intensidade do campo magnético está em função de uma corrente elétrica.



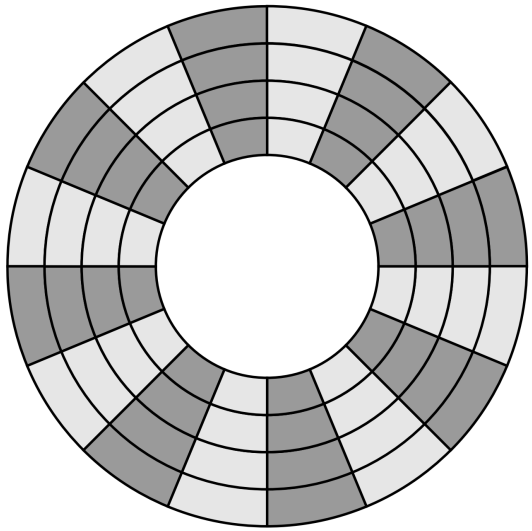
Disco magnético

- A cabeça L-G fica **MUITO** próxima do disco: 10 nm. Para referência: um fio de cabelo tem 75.000 nm de espessura; o vírus Influenza, 100 nm.
- Distância cabeça-disco de 40 nm, a densidade de bits é de 10 GB/pol^2. A 20 nm, 45 GB/pol^2;
- Dentro do disco tem ar (ou hélio) e o giro do disco movimenta o ar. A cabeça L-G sobrevoa o disco. Dependendo da atmosfera, pode aumentar a quantidade de discos dentro de um HD;
- Na figura anterior: LMR (*Longitudinal Magnetic Recording*). Existem o PMR (*Perpendicular*), SMR (*Shingled*) e o HAMR (*Heat-Assisted*) com maiores densidades de bits.

O braço atuador baseia-se na Força de Lorentz para se movimentar com desgaste mínimo.

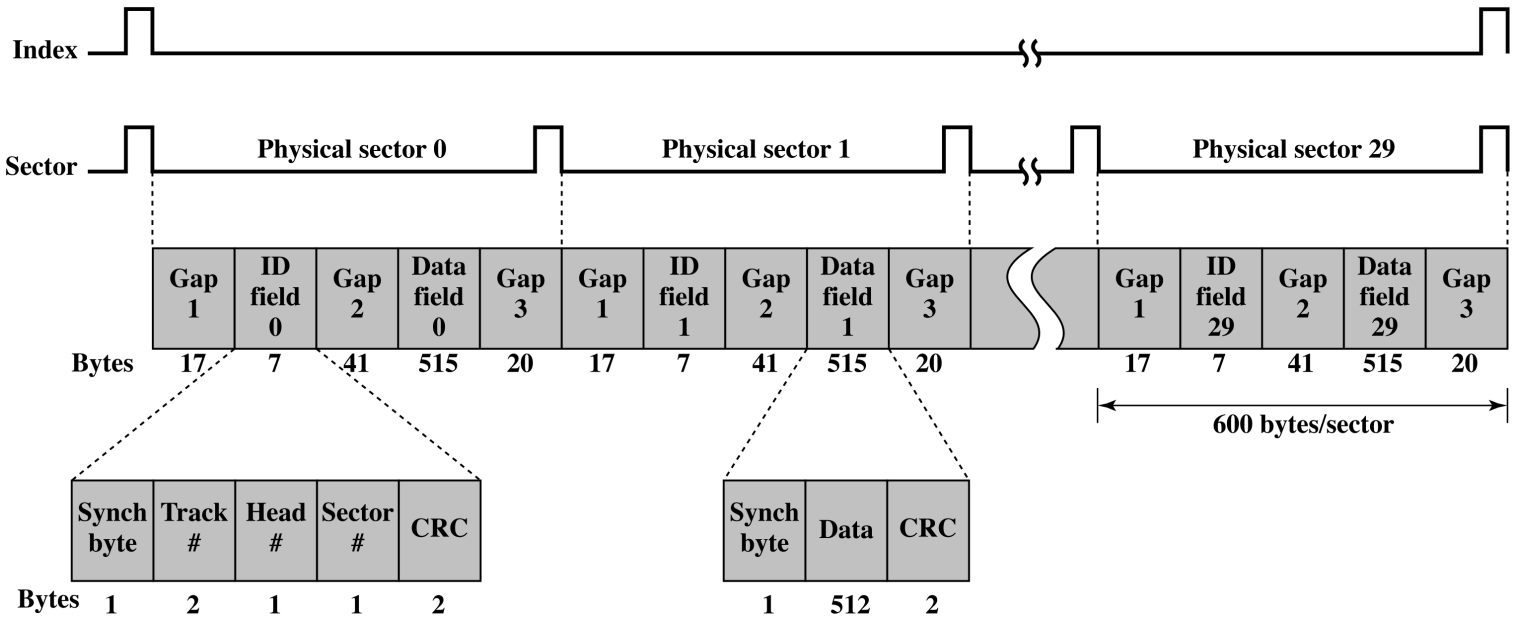
Disco magnético

- Um *layout* para disco magnético é chamado CAV (*Constant Angular Velocity*);
- No CAV, o disco gira a uma velocidade *angular* constante;
- Os blocos podem ser endereçados diretamente por trilha e setor;
- Considerando CAV, a velocidade *linear* nas extremidades é maior do que no centro;
- A densidade de bits/cm^2 é limitada pelas trilhas no interior do disco e diminui à medida que se afasta do centro (assim como a velocidade);



Disco magnético

- Dado que a cabeça L-G está na trilha correta, como o HD encontra o setor desejado?
- O disco possui um formato específico que indica cada setor;
- HDs usavam setores de 512 bytes. Hoje utilizam setores de 4096 bytes (*Advanced Format*) para maior eficiência.



Sync byte: byte de sincronização com um padrão específico de bits para indicar o início do setor.

Disco magnético – características físicas

Quanto ao movimento da cabeça de L-G:

- **Cabeças L-G fixas** – uma para cada trilha – montadas em um braço rígido e fixo (raro);
- Uma **cabeça L-G móvel** para a superfície toda e é montada em um braço móvel.

Quanto à portabilidade:

- **Discos não-removíveis**: são montados permanentemente dentro do dispositivo;
- **Discos removíveis**: o disco magnético pode ser removido e substituído por outro na unidade.

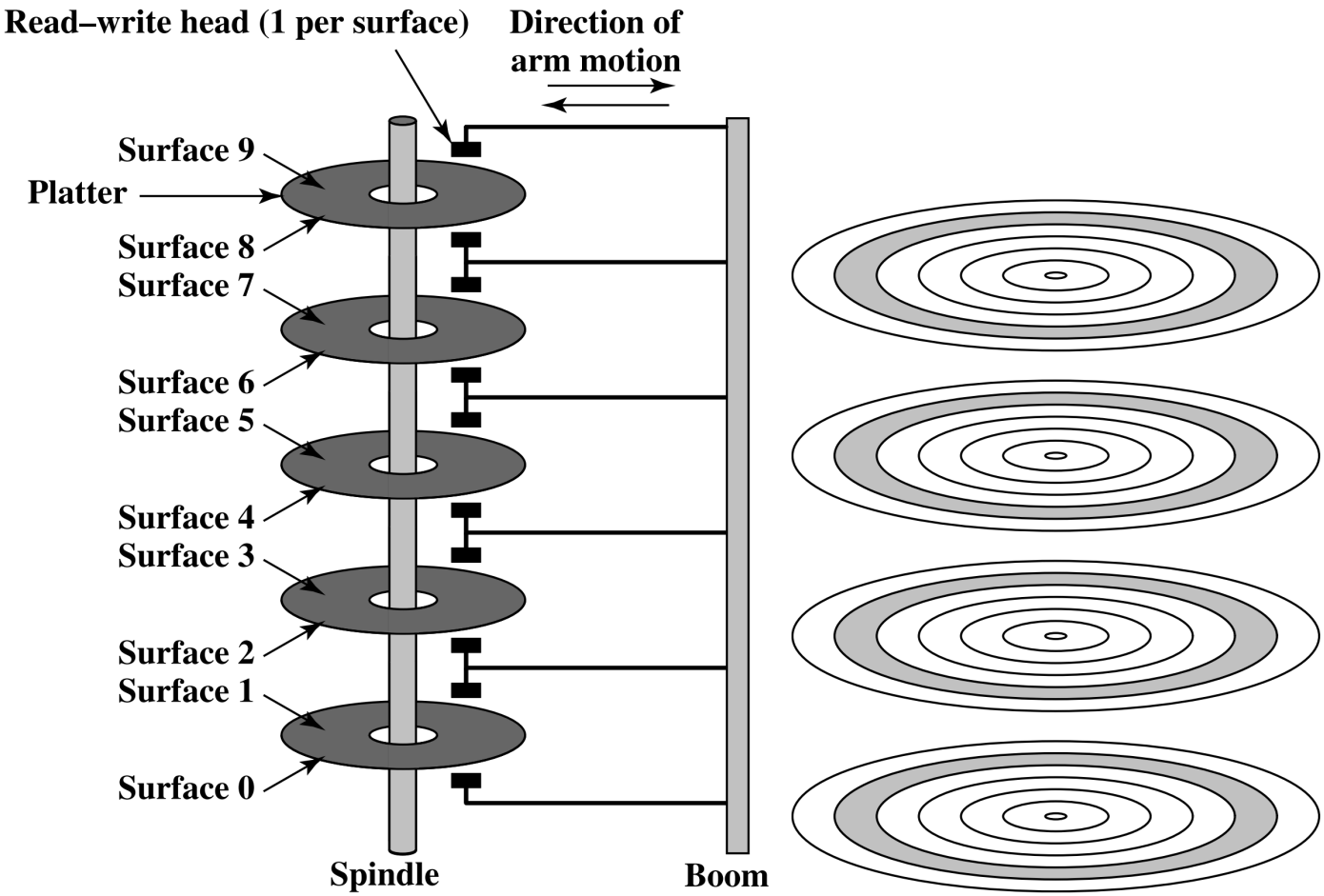
Quanto aos lados:

- **Single-sided**: apenas um dos lados do disco é utilizado;
- **Double-sided**: os dois lados do disco são utilizados.

Disco magnético – características físicas

Quando ao número de discos:

- "**Único prato**";
- "**Múltiplos pratos**" – conceito de **cilindro**.



Disco magnético – características físicas

Quando ao posicionamento das cabeças L-G:

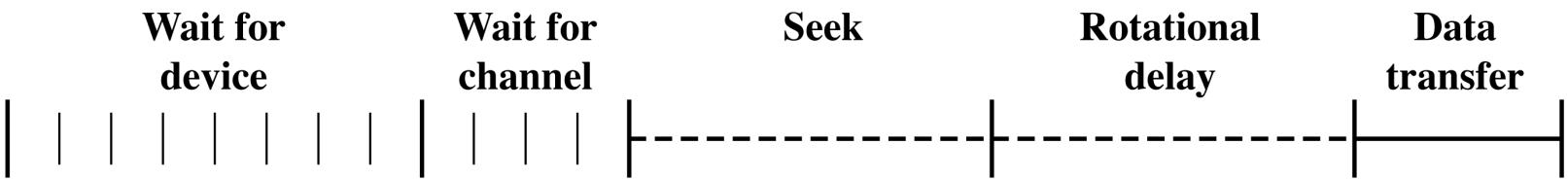
- **Posição fixa** acima do disco magnético, com uma separação de ar;
- **Em contato** com o disco magnético, como nas unidades removíveis e flexíveis (aka *floppy disks* ou disquetes);
- **Posição aerodinâmica**, sobrevoando bem próximo ao disco magnético, para maior densidade de bits, mas o disco pode ter imperfeições – *head crash*. É o caso dos HDs.

Disco magnético – parâmetros de desempenho

Um disco magnético possui desempenho definido por:

- Se ele possuir cabeça L-G móvel: **tempo de busca** (*seek time*): tempo para movimentar a cabeça L-G para a trilha desejada;
- **Latência (ou atraso) rotacional** (*rotational latency or delay*): tempo para o disco girar até que o setor alcance a cabeça L-G;
- **Tempo de acesso**: Tempo de busca + Latência rotacional;
- **Tempo de transferência**: tempo necessário para transferir (ler ou gravar) os dados;

Além disso, há tempos relativos à espera da disponibilidade do canal e do dispositivo.



Disco magnético – parâmetros de desempenho

Discos magnéticos...

- possuíam diâmetro de 14 polegadas, hoje variam entre 2,5" e 3,5" – menor distância para o braço percorrer. O tempo de médio busca T_b atualmente é menor do que 0,010 s;
- rotacionavam a uma velocidade de 3.600 rpm, mas podem chegar a 20.000 rpm. HDs de notebooks e de desktops variam entre 5.400 rpm e 7.200 rpm. HDs empresariais entre 10 mil e 15 mil rpm. Tempo de rotação:

$$T_r = \left(\frac{\text{RPM}}{60}\right)^{-1} \text{ [s]}$$

- Tempo médio de rotação: $T_{mr} = T_r \div 2$ [s];
- Tempo de transferência de b bytes em um disco com trilhas de N bytes:

$$T_t = \frac{b}{\frac{\text{RPM}}{60} \times N} \text{ [s]}$$

- Tempo médio de acesso: $T_a = T_b + T_{mr} + T_t$ [s].

Disco magnético – parâmetros de desempenho

Exemplo: considere um disco com tempo médio de busca de 0,004 s, velocidade de rotação de 15.000 rpm, setores de 512 bytes e 500 setores por trilha. Deseja-se transferir um bloco de 2.500 setores *contíguos* (ou 1.280.000 bytes). Qual o tempo de transferência?

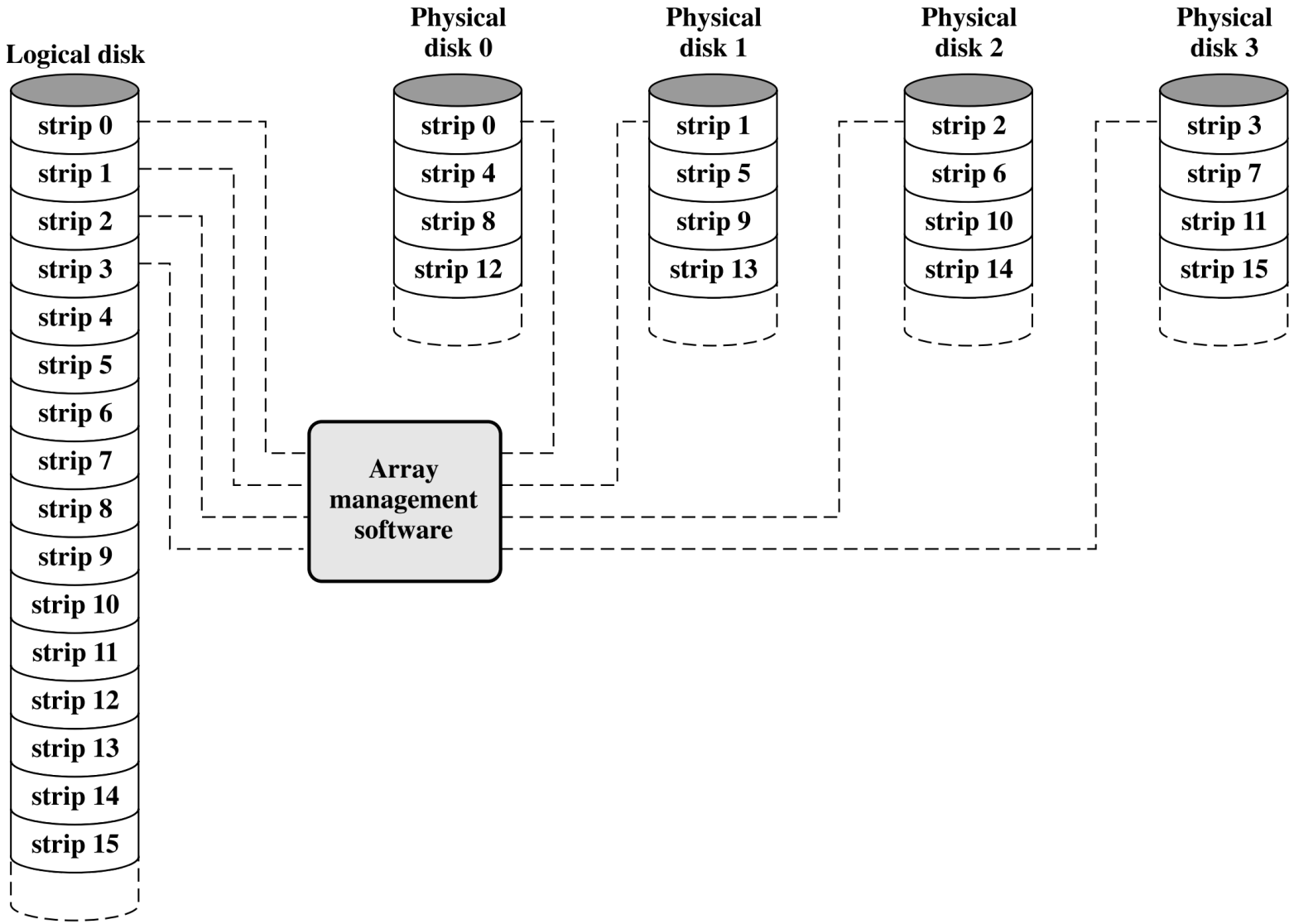
- 2.500 setores $\div$ 500 setores/trilha $=$ 5 trilhas *adjacentes* (considere nenhum tempo de busca entre uma trilha e outra);
- Tempo de rotação: $T_r = \left(\frac{15000}{60}\right)^{-1} = (250)^{-1} = 0{,}004$ s;
- Tempo médio de rotação: $T_{mr} = T_r \div 2 = 0{,}002$ s;
- Tempo para transferir 500 setores: $T_t = \frac{500 \times 512}{\frac{15000}{60} \times (500 \times 512)} = \frac{256000}{250 \times 256000} = 0{,}004$ s
- Tempo para transferir os primeiros 500 setores: $T_{a1} = 4 + 2 + 4 = 0{,}010$ s;
- Tempo para transferir outro bloco de 500 setores na trilha adjacente: $T_{a(2,5)} = 0{,}002 + 0{,}004$ s;
- Tempo total para transferir o arquivo: $T_T = T_{a1} + 4 \times T_{a(2,5)} = 0{,}010 + 4 \times 0{,}006 = 0,010 + 0,024 = 0,034$ s.

E no caso arquivo estar fragmentado em blocos de 500 setores?
E no caso de SSD?

RAID

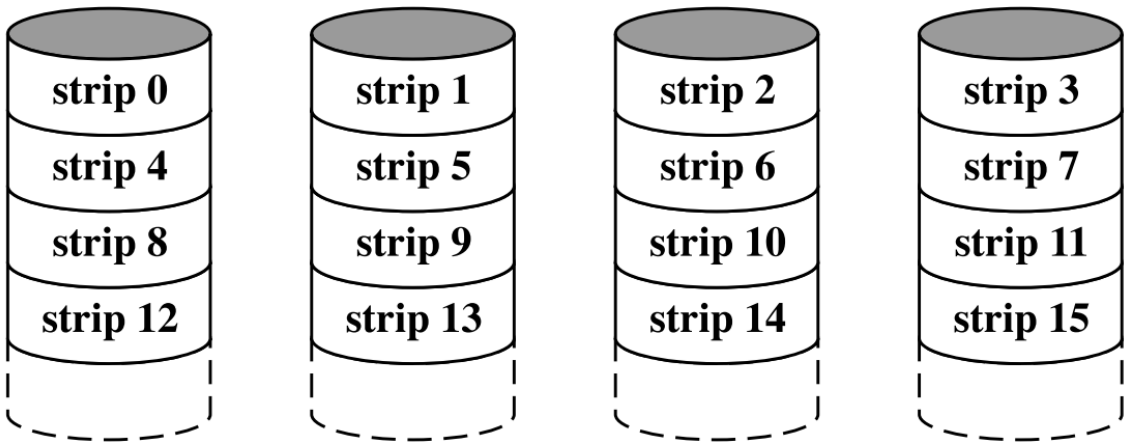
- Evolução no desempenho dos HDs é consideravelmente mais lenta do que da CPU e RAM == gargalo!
- Alternativa para aumentar o *throughput* $\rightarrow$ usar HDs independentes em paralelo;
- Múltiplas operações de I/O podem ser realizadas em paralelo, desde que os dados estejam em discos distintos;
- Uma única operação de I/O pode ser realizada em paralelo se os dados estiverem distribuídos em diversos discos;
- Pode-se aumentar a **confiabilidade** ao adicionar **redundância**;
- Para isso existe o RAID (*Redundant Array of Independent Disks*):
  - RAID propõe atender a necessidade por redundância;
  - Um arranjo RAID é visto pelo SO como um dispositivo lógico;
  - Dados são distribuídos pelos discos – *striping*;
  - Vários HDs atuando *aumentam* a probabilidade de falha;
  - Dados redundantes são armazenados para garantir a recuperabilidade em caso de falha de disco no arranjo.

RAID


Logical disk
strip 0
strip 1
strip 2
strip 3
strip 4
strip 5
strip 6
strip 7
strip 8
strip 9
strip 10
strip 11
strip 12
strip 13
strip 14
strip 15
Array management software
Physical disk 0
strip 0
strip 4
strip 8
strip 12
Physical disk 1
strip 1
strip 5
strip 9
strip 13
Physical disk 2
strip 2
strip 6
strip 10
strip 14
Physical disk 3
strip 3
strip 7
strip 11
strip 15

RAID - Níveis de RAID

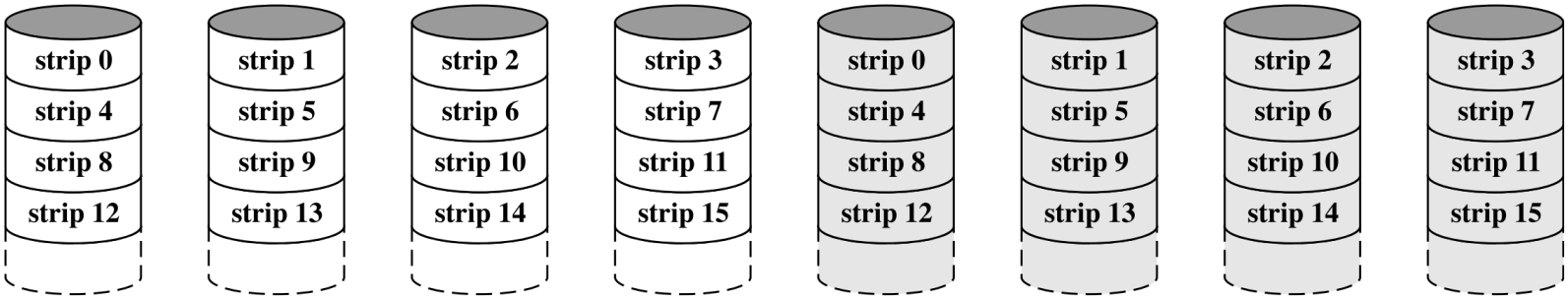
RAID 0/*striping*: **não** inclui redundância e possui altos desempenho (*throughput* e latência) e capacidade. Dados distribuídos (*striped*) pelos N discos do arranjo. Todos os N discos estão disponíveis para armazenar dados.



(a) RAID 0 (Nonredundant)

RAID - Níveis de RAID

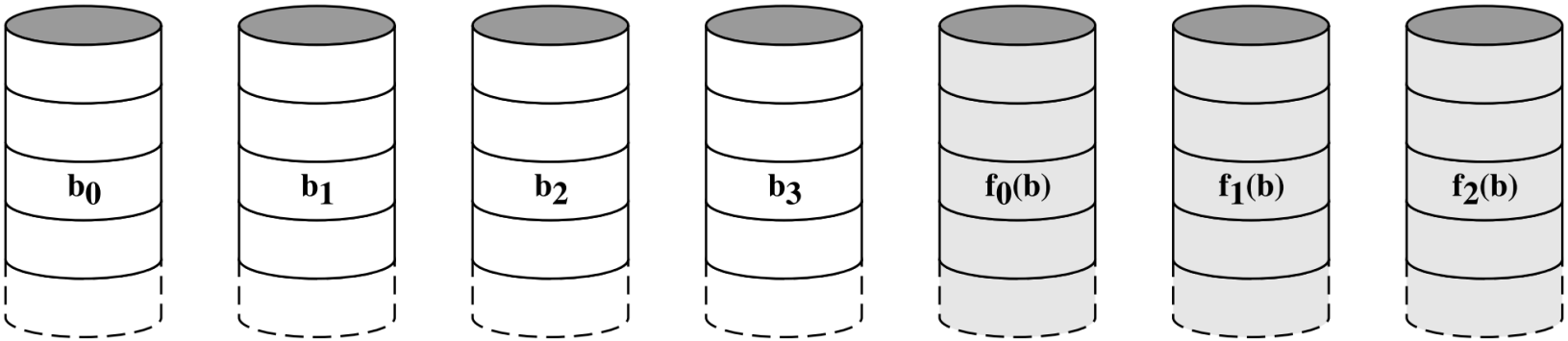
RAID 1/***mirroring***: Os dados são duplicados pelos N discos do arranjo. A capacidade do arranjo é de apenas **um** disco.A recuperação de dados é fácil, mas o RAID 1 é muito caro (utiliza $2N$ discos).



(b) RAID 1 (Mirrored)

RAID - Níveis de RAID

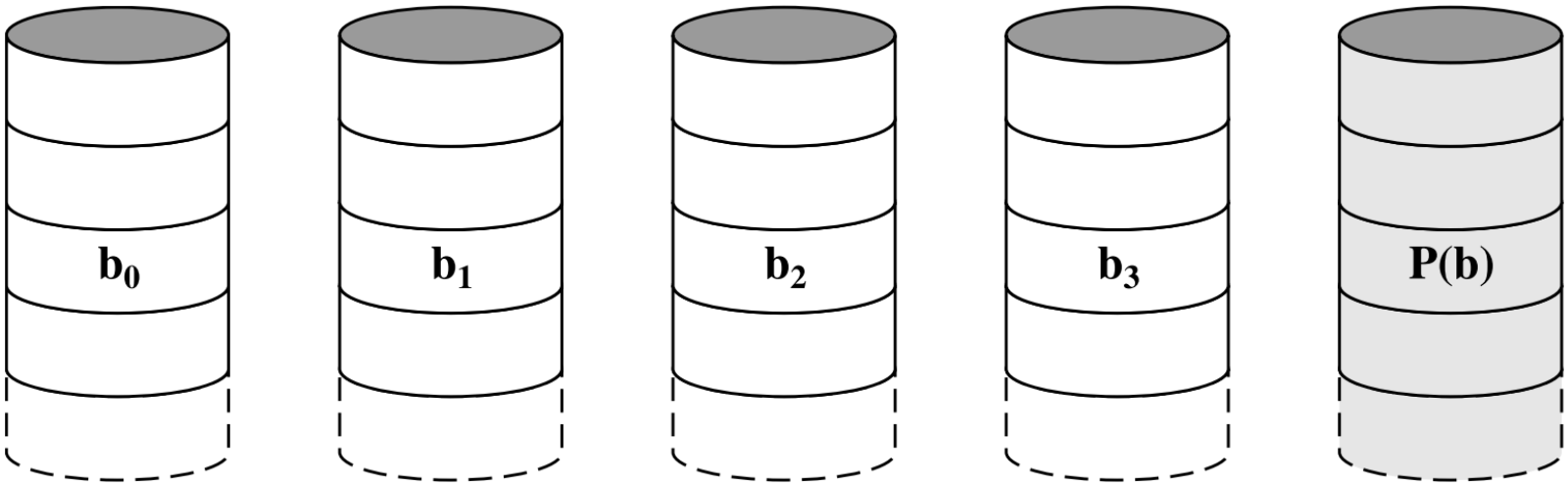
RAID 2/acesso paralelo com código de Hamming: Os discos estão sincronizados e todos eles participam de todas as operações de I/O. As *strips* são bem pequenas (byte ou *word*). Utiliza $N + m$ discos.



(c) RAID 2 (Redundancy through Hamming code)

RAID - Níveis de RAID

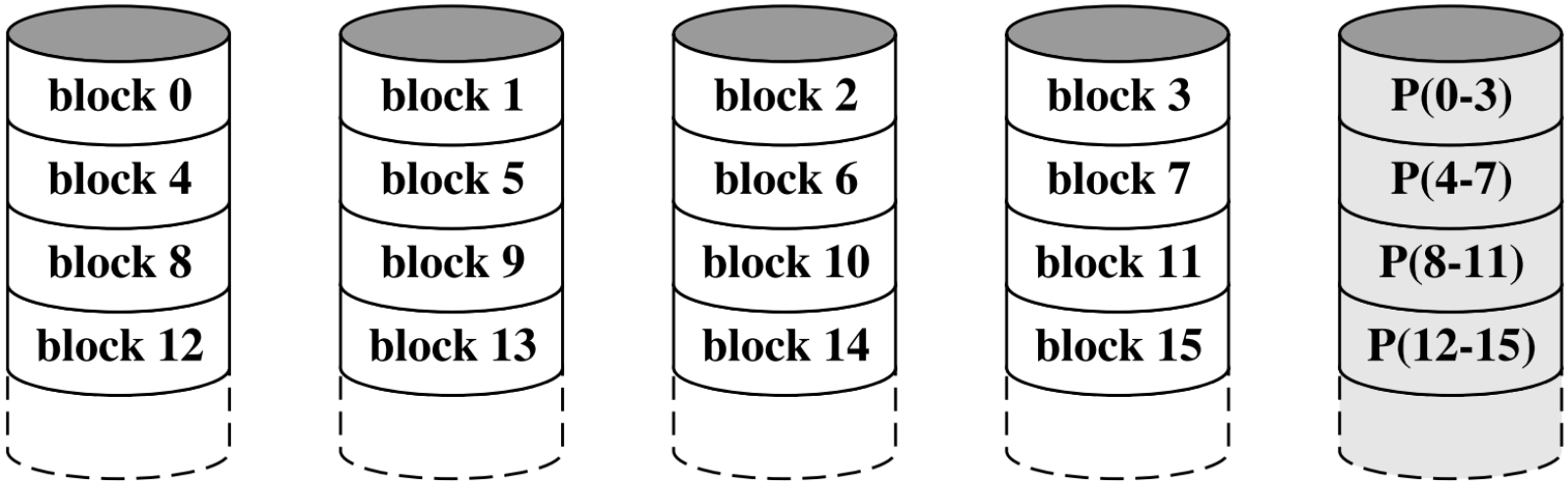
RAID 3/acesso paralelo com paridade: Similar ao RAID 2, mas utiliza paridade (XOR) em vez do código de Hamming. Utiliza $N + 1$ discos.



(d) RAID 3 (Bit-interleaved parity)

RAID - Níveis de RAID

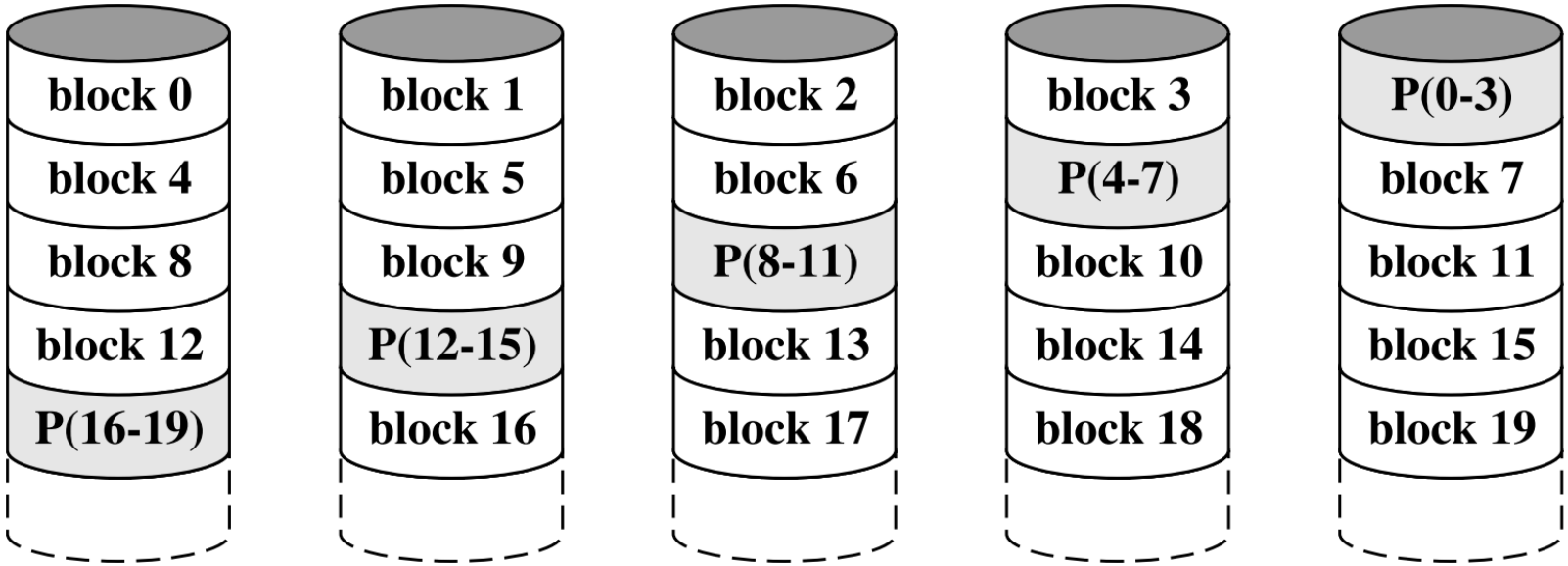
RAID 4/acesso independente: Os discos operam independentemente e operações de I/O podem ser atendidas em paralelo. As *strips* são maiores e todas as gravações exigem a atualização da paridade – o disco de paridade pode se tornar um gargalo (além de sofrer mais desgaste por uso) e deve ler os dados dos outros discos. Utiliza $N + 1$ discos.



(e) RAID 4 (Block-level parity)

RAID - Níveis de RAID

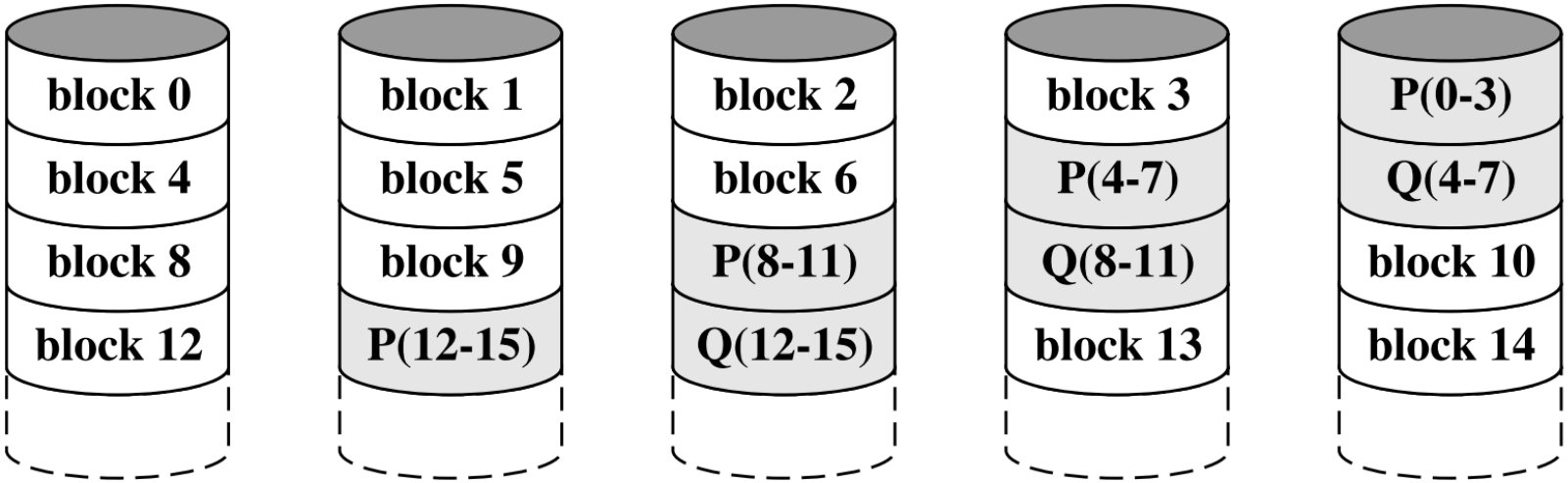
RAID 5/acesso independente com paridade distribuída: semelhante ao RAID 4, mas distribui a paridade dos blocos por todos os $N + 1$ discos.



(f) RAID 5 (Block-level distributed parity)

RAID - Níveis de RAID

RAID 6/acesso independente com paridade dupla distribuída: semelhante ao RAID 5, mas utiliza dois algoritmos de checagem de dados: a paridade dos RAID 4/5 e outro independente. Escritas são extremamente custosas. Utiliza $N + 2$ discos.



(g) RAID 6 (Dual redundancy)

RAID

Category	Level	Description	Disks Required	Data Availability	Large I/O Data Transfer Capacity	Small I/O Request Rate
Striping	0	Nonredundant	N	Lower than single disk	Very high	Very high for both read and write
Mirroring	1	Mirrored	$2N$	Higher than RAID 2, 3, 4, or 5; lower than RAID 6	Higher than single disk for read; similar to single disk for write	Up to twice that of a single disk for read; similar to single disk for write
Parallel access	2	Redundant via Hamming code	$N + m$	Much higher than single disk; comparable to RAID 3, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
	3	Bit-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
Independent access	4	Block-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 5	Similar to RAID 0 for read; significantly lower than single disk for write	Similar to RAID 0 for read; significantly lower than single disk for write
	5	Block-interleaved distributed parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 4	Similar to RAID 0 for read; lower than single disk for write	Similar to RAID 0 for read; generally lower than single disk for write
	6	Block-interleaved dual distributed parity	$N + 2$	Highest of all listed alternatives	Similar to RAID 0 for read; lower than RAID 5 for write	Similar to RAID 0 for read; significantly lower than RAID 5 for write

N = number of data disks; m proportional to log N

RAID

Level	Advantages	Disadvantages	Applications
0	I/O performance is greatly improved by spreading the I/O load across many channels and drives No parity calculation overhead is involved Very simple design Easy to implement	The failure of just one drive will result in all data in an array being lost	Video production and editing Image Editing Pre-press applications Any application requiring high bandwidth
1	100% redundancy of data means no rebuild is necessary in case of a disk failure, just a copy to the replacement disk Under certain circumstances, RAID 1 can sustain multiple simultaneous drive failures Simplest RAID storage subsystem design	Highest disk overhead of all RAID types (100%)—inefficient	Accounting Payroll Financial Any application requiring very high availability

RAID

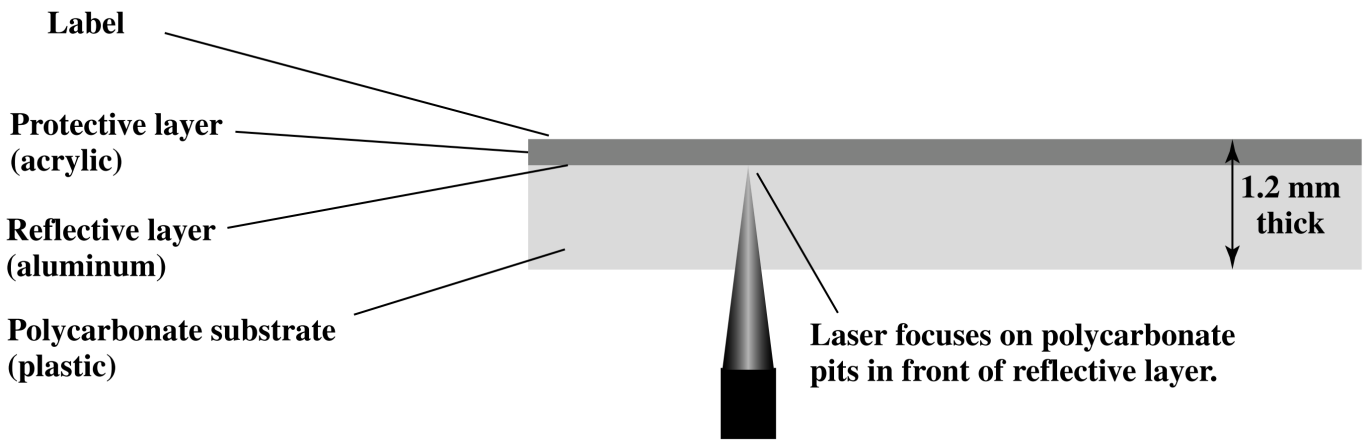
Level	Advantages	Disadvantages	Applications
2	Extremely high data transfer rates possible The higher the data transfer rate required, the better the ratio of data disks to ECC disks Relatively simple controller design compared to RAID levels 3, 4 & 5	Very high ratio of ECC disks to data disks with smaller word sizes—inefficient Entry level cost very high—requires very high transfer rate requirement to justify	No commercial implementations exist/ not commercially viable
3	Very high read data transfer rate Very high write data transfer rate Disk failure has an insignificant impact on throughput Low ratio of ECC (parity) disks to data disks means high efficiency	Transaction rate equal to that of a single disk drive at best (if spindles are synchronized) Controller design is fairly complex	Video production and live streaming Image editing Video editing Prepress applications Any application requiring high throughput

RAID

Level	Advantages	Disadvantages	Applications
4	Very high Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency	Quite complex controller design Worst write transaction rate and Write aggregate transfer rate Difficult and inefficient data rebuild in the event of disk failure	No commercial implementations exist/ not commercially viable
5	Highest Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency Good aggregate transfer rate	Most complex controller design Difficult to rebuild in the event of a disk failure (as compared to RAID level 1)	File and application servers Database servers Web, e-mail, and news servers Intranet servers Most versatile RAID level
6	Provides for an extremely high data fault tolerance and can sustain multiple simultaneous drive failures	More complex controller design Controller overhead to compute parity addresses is extremely high	Perfect solution for mission critical applications

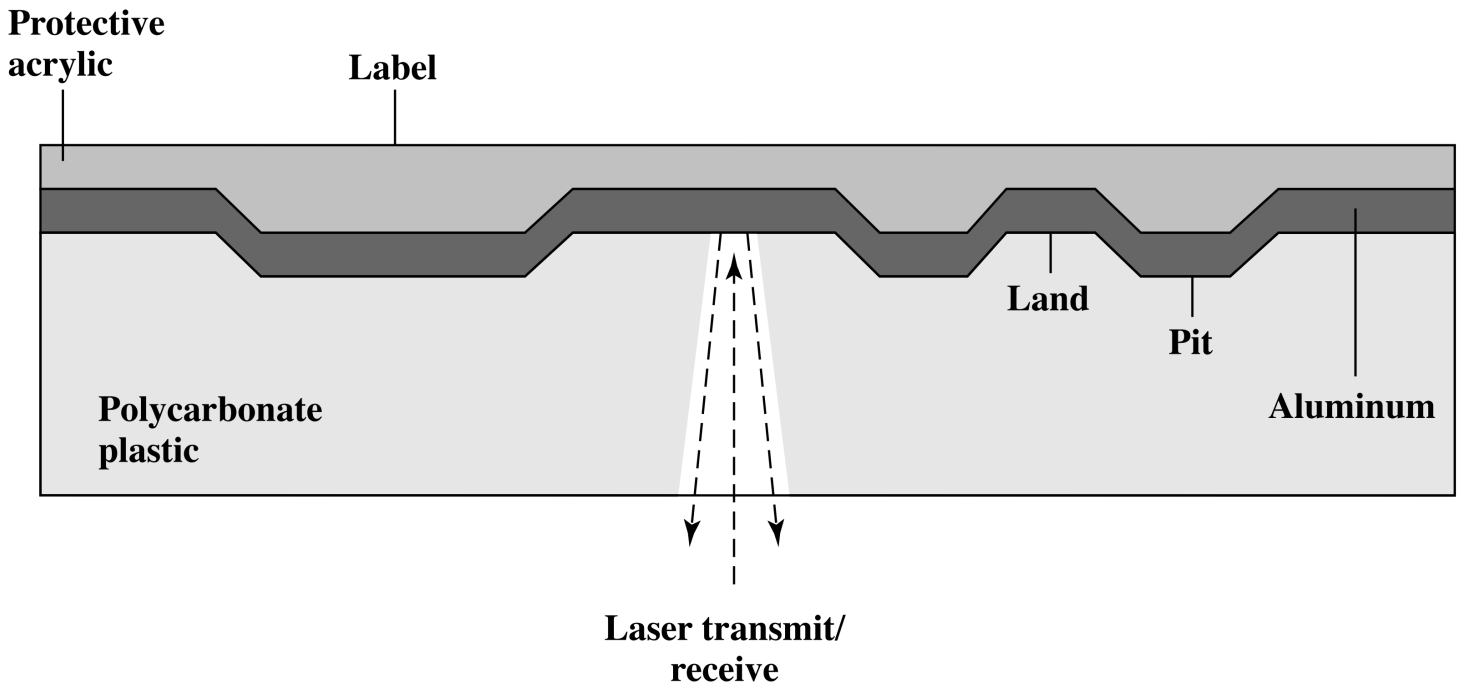
Discos ópticos

- O CD (*Compact Disk*) foi introduzido em 1983;
- Mídia não-apagável e armazenava até 60 minutos de áudio digital no formato PCM 44,1 KHz, estéreo, 16 bits;
- CD-ROM (700 MB/80 min) utiliza técnicas de ECC para garantir a transferência correta de dados de computador;
- É construído da seguinte maneira:



Discos ópticos

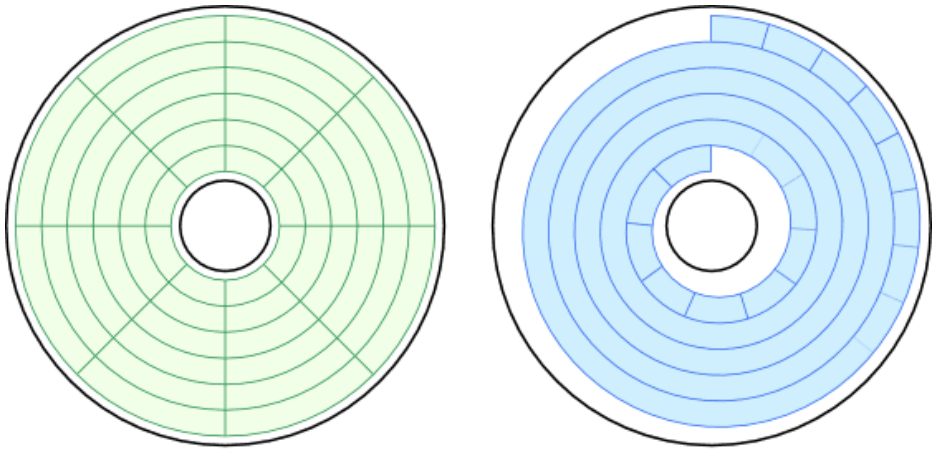
- O CD-ROM e outras mídias ópticas em geral funcionam da seguinte maneira:



- Utiliza-se um laser de baixa potência;
- *Lands* é uma superfície suave e reflete o *laser* com mais intensidade. As covas (*pits*) são ásperas e menos reflexivas;

Discos ópticos

- Para maior eficiência de capacidade, os dados são armazenados em uma única trilha em espiral de dentro para fora do disco;

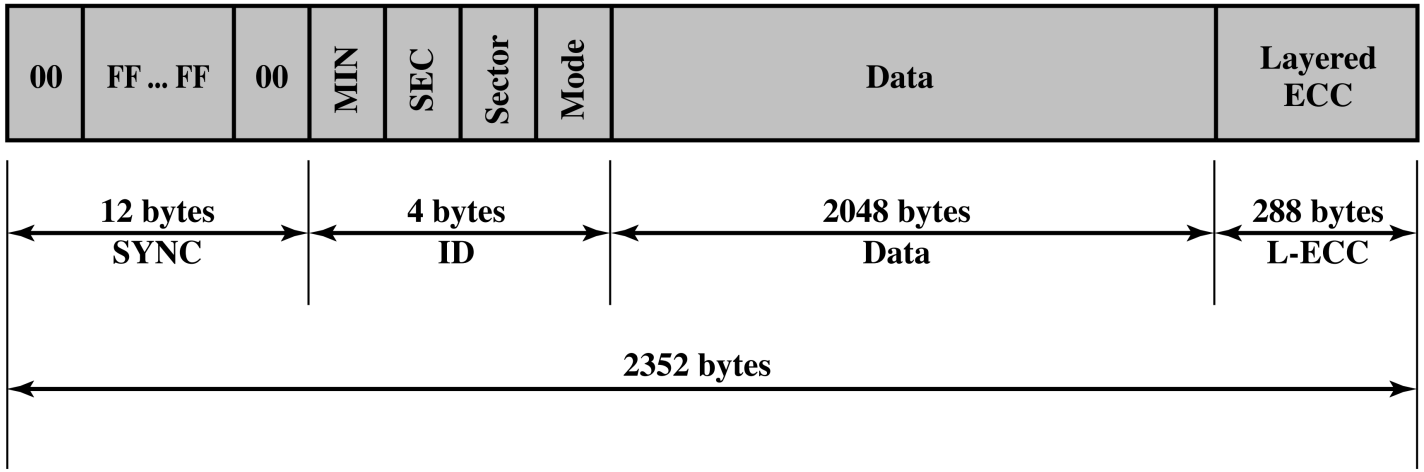


Constant angular velocity disc | Constant linear velocity disc

- O disco gira a velocidades angulares distintas para manter a mesma velocidade *linear* – CLV (*Constant Linear Velocity*);
- A busca de setores é mais difícil – ↑o tempo de busca para operações aleatórias.

Discos ópticos

- O CD-ROM é formatado da seguinte maneira:



- **Sync**: indica o início de um bloco – oito bits 0, 10 bytes 1, oito bits 0;
- **Header**: indica o endereço do bloco e o modo de armazenamento: modo 1: 2048 bytes + ECC; modo 2: 2336 bytes sem ECC;
- **Data**: dados de usuário;
- **Auxiliary**: 228 bytes para ECC (modo 1) ou dados de usuário (modo 2).

Discos ópticos

- O CD-ROM é apropriado para a distribuição de grandes volumes de dados para muitos usuários;
- O processo de gravação inicial é caro, mas as cópias subsequentes são baratas;
- É removível, permitindo ser usado em arquivamento de dados;
- Mas não pode ser regravado e o tempo de acesso é muito maior, até 0,5 s.

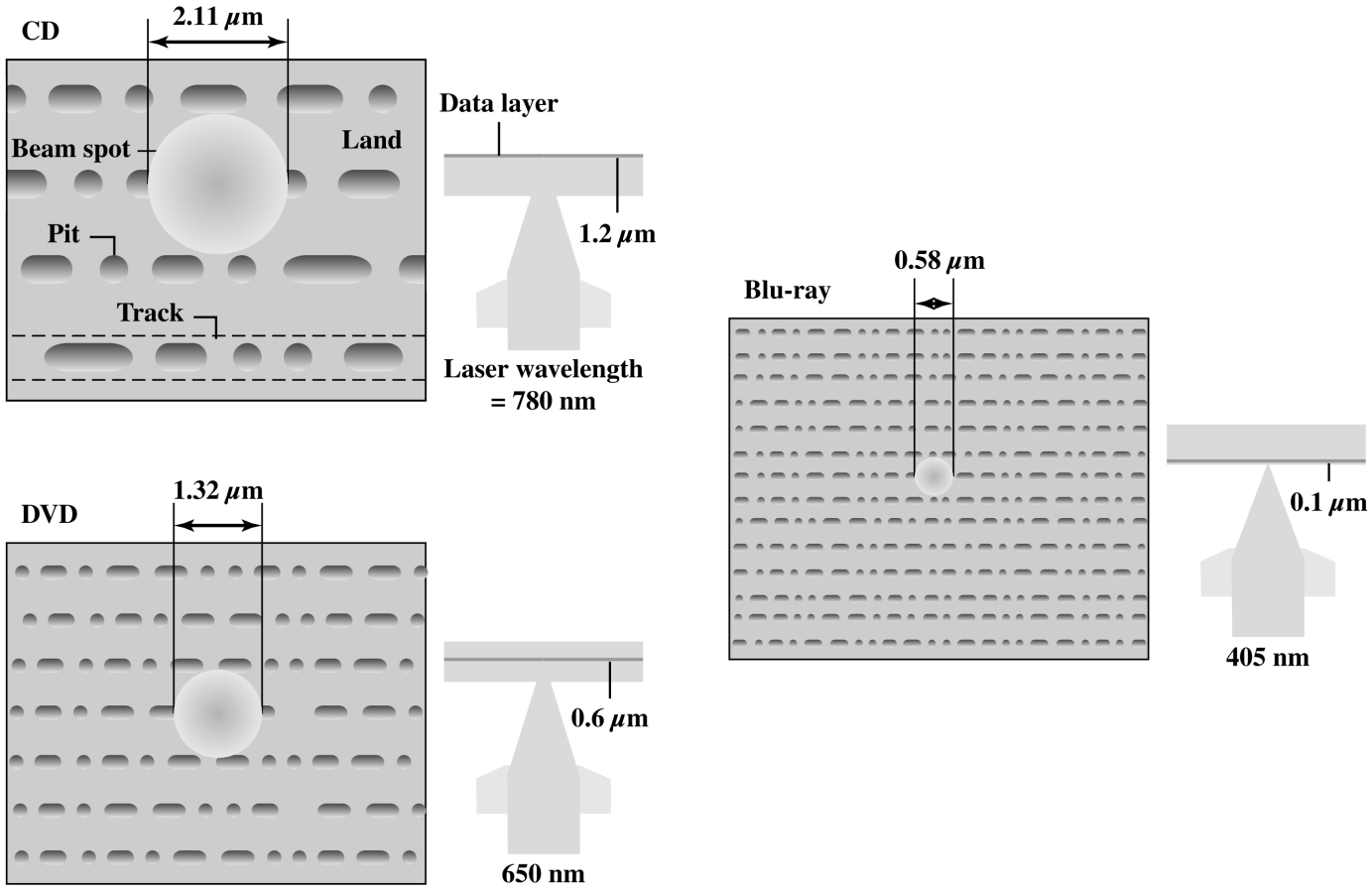
Discos ópticos

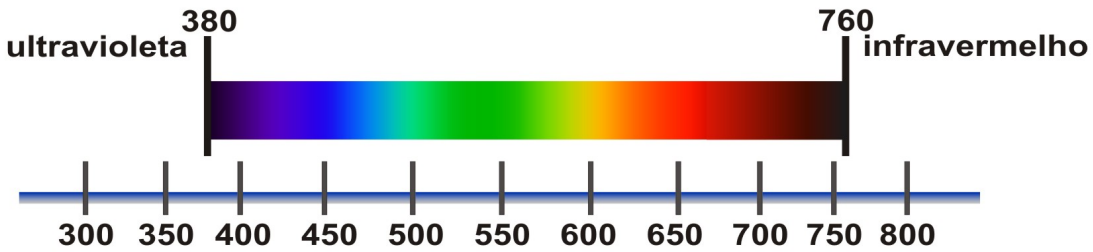
- O CD-R (*Recordable*) é parecido com o CD-ROM e é apropriado para a gravação de pequeno número de cópias;
- Utiliza um laser de intensidade média que altera a refletividade da mídia óptica ("queima o CD");
- Após gravado, o CD-R só pode ser lido;
- O CD-RW (*Rewritable*) pode ser reescrito diversas vezes;
- Utiliza a estratégia de **mudança de fase**: a mídia óptica pode se tornar amorfa ou cristalina de acordo com a intensidade do laser;
- Tem um número máximo de ciclos de apagamento: entre 500 mil e um milhão;

Discos ópticos

- DVD-ROM: com capacidade maior, devido à maior densidade de bits, substituiu o VHS e é utilizado para vídeos;
- Pode armazenar entre 4,7 GB (*single layer*) e 8,5 GB (*dual layer*, com ajuste de foco do laser) em cada lado;
- Usando os dois lados e *dual-layer*: 17 GB;
- Além disso: DVD-R e DVD-RW.
- Blu-ray (BD-ROM): utilizado na distribuição de filmes em alta definição;
- Capacidades variando entre 25 GB (*single layer*) e 50 GB (*dual layer*);
- Também: BD-R e BD-RE (*Recordable Erasable*);
- Ultra HD Blu-ray: capacidades variando entre 50 GB (*single layer*), 66 GB (*dual layer*), 100 GB (*triple layer*).

Discos ópticos



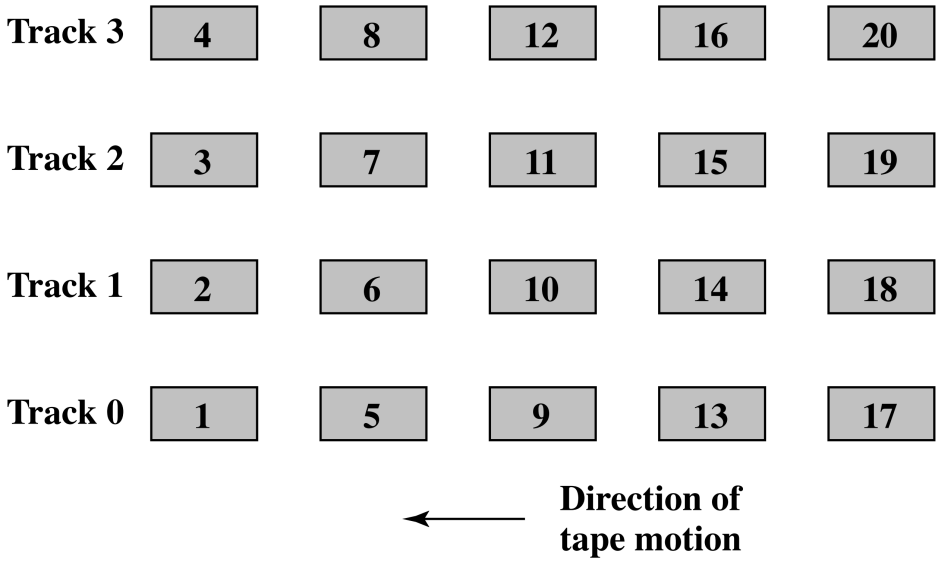


Fita magnética

- Fitas magnéticas utilizam a mesma técnica de leitura/gravação dos discos magnéticos;
- Mas são um dispositivo de *acesso sequencial* (lembrando que HDs são de *acesso direto*);
- A mídia é uma fita de poliéster flexível coberta com material magnetizável e, atualmente, são encapsuladas em cartuchos;
- Fitas foram o primeiro tipo de memória externa;
- Hoje são utilizadas como o último integrante na hierarquia de memória.

Fita magnética

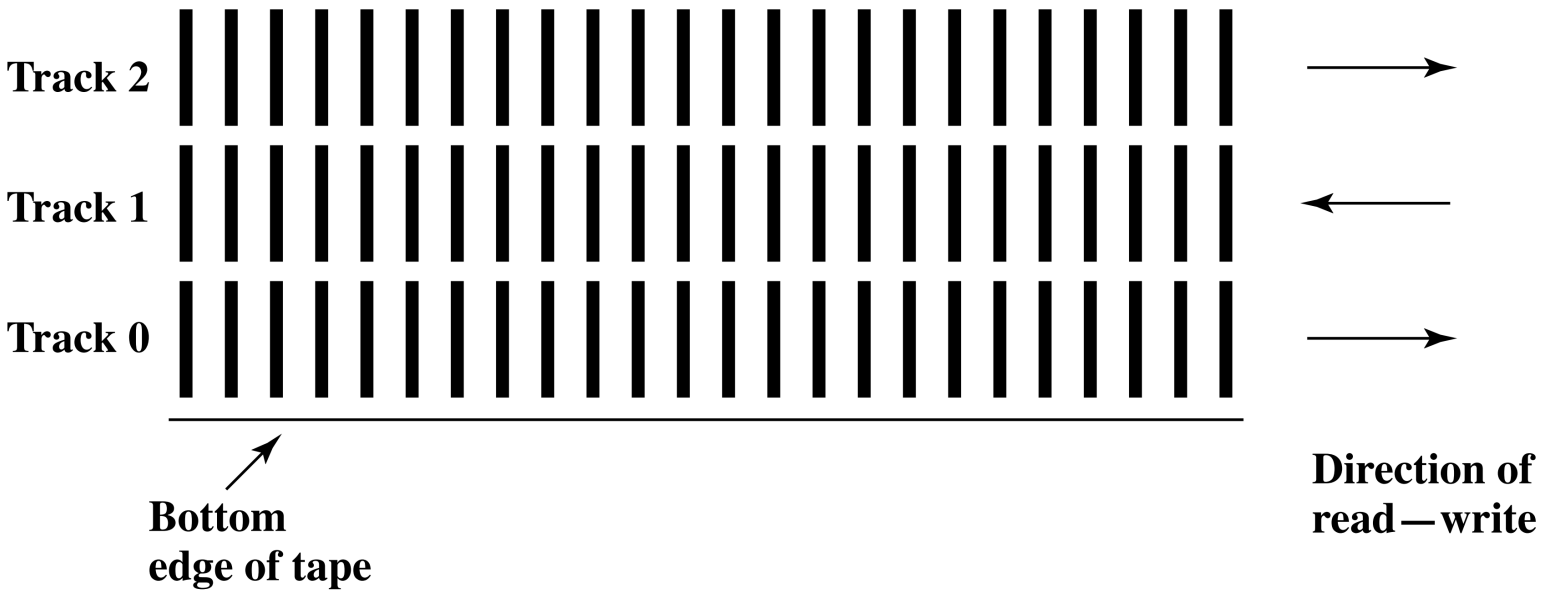
- Os dados podem ser organizados em trilhas **paralelas** no sentido longitudinal. Tecnologias antigas possuíam nove trilhas (8 bits + paridade), depois passaram a utilizar 18 (16 bits + checagem) ou 36 (32 bits + checagem) trilhas:



(b) Block layout for system that reads—writes four tracks simultaneously

Fita magnética

- Entretanto, tecnologias atuais utilizam gravação **serial** ao longo das trilhas;
- Blocos (aka *physical records*) são separados por *interrecord gaps*;
- Além disso, utiliza-se a gravação em **serpentina**:



(a) Serpentine reading and writing

Fita magnética

- A tecnologia de fita utilizada atualmente é a LTO (*Linear Tape-Open*):

	LTO-1	**LTO-2**	**LTO-3**	**LTO-4**	**LTO-5**	**LTO-6**	**LTO-7**	**LTO-8**
Release date	2000	2003	2005	2007	2010	TBA	TBA	TBA
Compressed capacity	200 GB	400 GB	800 GB	1600 GB	3.2 TB	8 TB	16 TB	32 TB
Compressed transfer rate	40 MB/s	80 MB/s	160 MB/s	240 MB/s	280 MB/s	525 MB/s	788 MB/s	1.18 GB/s
Linear density (bits/mm)	4880	7398	9638	13250	15142			
Tape tracks	384	512	704	896	1280			
Tape length (m)	609	609	680	820	846			
Tape width (cm)	1.27	1.27	1.27	1.27	1.27			
Write elements	8	8	16	16	16			
WORM?	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Encryption Capable?	No	No	No	Yes	Yes	Yes	Yes	Yes
Partitioning?	No	No	No	No	Yes	Yes	Yes	Yes

SSD

- SSDs infelizmente não são cobertos na 8ª edição, só na 9ª edição em inglês;
- Material disponível online;
- Mas apenas cobre um pouco de *Flash-based SSDs*, não fala nada sobre *3D Xpoint-based SSDs*, nem sobre a evolução dos SSDs.

Exercícios

Capítulo abordado: 6